General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

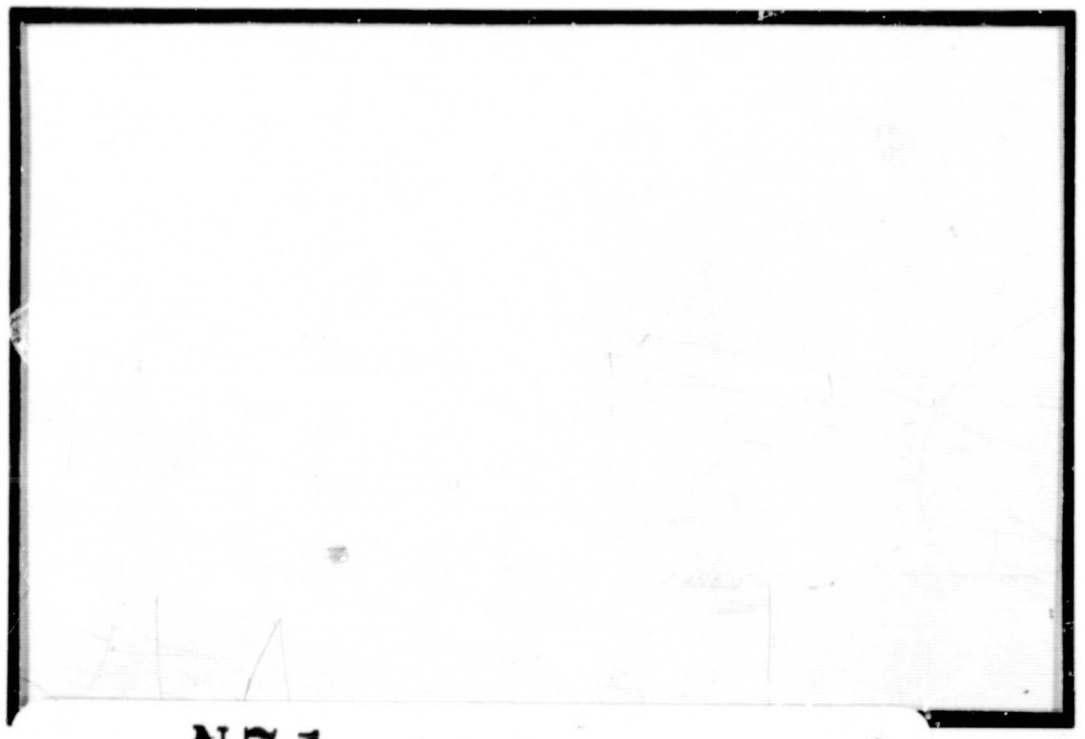
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

ELECTRICAL ENGINEERING



ENGINEERING EXPERIMENT STATION

AUBURN UNIVERSITY

AUBURN, ALABAMA

19

ERROR ANALYSIS OF NUMERICAL INTEGRATION SCHEMES

PREPARED BY

GUIDANCE AND CONTROL STUDY GROUP

JOSEPH S. BOLAND, III, PROJECT LEADER

TWENTY THIRD TECHNICAL REPORT

NOVEMBER 15, 1970

CONTRACT NAS8-20104
GEORGE C. MARSHALL SPACE FLIGHT CENTER
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
HUNTSVILLE, ALABAMA

APPROVED BY:

Chester C. Carroll
Professor and Head
Electrical Engineering

SUBMITTED BY:

Joseph S. Boland, III
Assistant Professor
Electrical Engineering

FOREWORD

This report is a technical summary of the progress made by the Electrical Engineering Department, Auburn University, toward fulfillment of Contract NAS8-20104 granted to Auburn Research Foundation, Auburn, Alabama. This contract was awarded April 6, 1965, by the George C. Marshall Space Flight Center, National Aeronautics and Space Administration, Huntsville, Alabama.

A thesis to be submitted by Brij Bhushan to the Graduate Faculty of Auburn University in partial fulfillment of the requirements for the degree of Master of Science is based on the work reported herein.

SUMMARY

This study deals with the accuracy of numerical integration schemes and considers the computational errors involved therein. The round-off errors are assumed to be random in nature and simulations are run to investigate the effect of noisy inputs on the errors. Simulations are run to study the effect of the mode of operation and the change in input level on the errors. Some conclusions are drawn from the results of these simulations. A method based on the concept of "Practical Stability" is developed and shown to be applicable among other examples to the error equation developed from the differential equation to be integrated, thus giving the error bounds of this system.

PERSONNEL

The following staff members of Auburn University are active participants in the work of this contract:

J. S Boland, III - Assistant Professor of Electrical Engineering

Brij Bhushan - Graduate Research Assistant, Electrical Engineering

Michael H. Fong - Graduate Research Assistant, Electrical Engineering

D. W. Sutherlin - Graduate Research Assistant, Electrical Engineering

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GENERAL NOTATION

Let ε denote set membership, let R^n denote a real Euclidean n-space, let $||\cdot||$ denote a norm, let $\subset$ denote set inclusion and let $J = [t_o, \infty)$, $-\infty < t_o < \infty$. If A and B are sets, let AxB denote their cartesian product. Let ϕ denote the null set.

Systems are considered which may be represented by equations of the form

$$\frac{dx}{dt} = f(x,t) \tag{1}$$

where in (1) $x \varepsilon R^n$, $f: R^n \times J \to R^n$. It is assumed that f is continuous on $R^n \times J$. The solutions of (1) are denoted by $x(t; x_o, t_o)$ with $x(t_o; x_o, t_o) = x_o$. In general it is not required that $f(0, t) = 0$. Let $S(t) \subset R^n$ for all $t \varepsilon J$. Assume that S(t) is a connected open region. Let $\overline{S(t)}$ denote the closure of S(t) and let $\delta S(t)$ denote the boundary of $\overline{S(t)}$. Assume that S(t) is bounded for all $t \varepsilon J$, that $\lim_{t \to t_a} S(t)$ exists for all $t_a \varepsilon J$ and that $\lim_{t \to t_a} S(t) = S(t_a)$. In the sequel the symbol S(t) (with appropriate subscripts) is assumed to have the properties described above.

Let $[S(t) - S_o(t)] = \{x \varepsilon R^n: x \varepsilon S(t), x \notin S_o(t)\}$

Let $B(a) = \{x \varepsilon R^n: ||x|| < a\}$

$\overline{B(a)} = \{x \varepsilon R^n: ||x|| \leq a\}$

In what follows, the mappings V: $R^n \times J \longrightarrow R^1$ are utilized. Further it is assumed that V belongs to C^1(i.e., V has continuous first partial derivatives on $R^n \times J$). $\dot{V}_{(1)}$ denotes the total derivative of V with respect to t evaluated along solutions of (1).

$$\dot{V}_{(1)}(x, t) = \frac{\delta V(x, t)}{\delta t} + \text{grad } (V(x, t))f(x, t)$$

The following list gives the meaning and description of the most repetitively used symbols that conserve their significance throughout this study.

LIST OF SYMBOLS

$P_V(t)$	time measurement in the vehicle co-ordinate system
$P_U(t)$	measurement in the navigational frame
$B(t)$	direction cosine matrix
$B^T(t)$	transpose of $B(t)$
$E(t)$	four parameter vector
$\cdot$ or $\frac{d}{dt}$	denotes derivative with respect to time
$\Phi(t)$	a square matrix relating $\dot{E}(t)$ and $E(t)$
$E_1(t), E_2(t), E_3(t), E_4(t)$	components of vector $E(t)$
$\dot{\Phi}_x, \dot{\Phi}_y, \dot{\Phi}_z$	body rotational rates along x, y and z axis of vehicle
h	the integration step size
$S(t)$	time varying sets with proper subscripts
$V(x, t)$	a real valued function
$\chi(t)$	a real valued function
Sup	supremum value
Inf	infinimum value
$G(s)$	system transfer function
$r(t)$	skew symmetric matrix of body angular rates
B_{k+1}	discrete direction cosine matrix
$\hat{B}_{k+1}$	approximate direction cosine matrix
χ_k	state transition matrix between states k and k+1

LIST OF SYMBOLS (cont.)

$\hat{X}_k$	approximate state transition matrix between k and k+1
$\tilde{X}_k$	difference between approximate and exact state transition matrix
I	Identity matrix
E_n	error between the approximate and exact direction cosine matrix

I. INTRODUCTION

The subject of real time digital flight simulation is an important one that has gained considerable importance in the past few years. A digital simulation most often requires the numerical integration of the corresponding differential system. It is here that one comes across problems such as stability, accuracy and efficiency of the given numerical integration scheme, computation time required, memory required etc.

This study deals with the accuracy of the numerical integration scheme used and considers the computational errors involved therein. Chapter II describes the basic set of differential equations which are to be integrated and the errors involved in their numerical integration by a digital machine.

Chapter III deals with the effect of random noise on some of these errors and presents results of some of the simulations carried out on an IBM/360 Model 50 digital computer using S/360 CSMP language and floating point mode of arithmetic calculations.

Chapter IV describes the results obtained by using Fortran IV programming language and using fixed point mode of calculation. It compares some of the results thus obtained with the results of Chapter III.

Chapter V defines some stability definitions and gives some theorems with proofs which when applied yield a bound on the system

trajectories. These theorems are then applied to the error equation and a quantitative bound on the error obtained. Several examples are worked using this technique.

II. TYPES OF ERRORS IN NUMERICAL COMPUTATION BY MACHINE

A digital simulation most often requires the numerical integration of the corresponding differential system. In so doing one comes across the limitations of the machine and of the technique used for these numerical integrations. Broadly speaking the errors introduced can be classified into the categories of round-off errors and truncation errors.

The round-off error arises from the finite word length of the computing machine. The local round-off error depends on many factors, such as: the number system employed by the machine, the machine word length, the mode of operation (fixed point or floating point), the sequence in which the numerical operations are arranged, and many others. Although for known environments, viz. all the factors mentioned above, the round off error should be a well defined quantity, its estimation is a complex process. It is therefore reasonable to assume that round off errors arise in an essentially unpredictable or random manner.

Truncation errors are caused by the type of numerical integration used and are deterministic in nature. The error bounds can be developed for a known integration scheme.

The space vehicle under consideration has a strapped-down inertial navigation system. The measurements of the vehicle rotational rates are made in the vehicle co-ordinate system. Since the transformation

of these measurements into the navigational co-ordinate system is dependent upon the instantaneous orientation of the vehicle, a co-ordinate transformation matrix B(t) must be calculated in order that the measurements in the vehicle co-ordinate system may be resolved into the navigational co-ordinate system. The transformation is given by:

$$P_n(t) = B^T(t)P_v(t) \tag{2}$$

where,

$P_v(t)$ = measurement in vehicle co-ordinate system.

$P_n(t)$ = measurements converted into the navigational frame.

$B(t)$ = Direction cosine matrix.

$B^T(t)$ = Transpose of matrix $B(t)$.

The direction cosine matrix can be easily calculated from the four parameter vector $E(t)$. For details refer to Burdeshaw [1]. The time rate of change of $E(t)$ using the four parameter method [1] is given by:

$$\dot{E}(t) = \Phi(t)\, E(t) \tag{3}$$

where

$$\dot{E}(t) = [\dot{E}_1(t),\ \dot{E}_2(t),\ \dot{E}_3(t),\ \dot{E}_4(t)]^T$$

$$\Phi(t) = \frac{1}{2}\begin{bmatrix} 0 & -\dot{\phi}_z & -\dot{\phi}_y & -\dot{\phi}_x \\ \dot{\phi}_z & 0 & -\dot{\phi}_x & \dot{\phi}_y \\ \dot{\phi}_y & \dot{\phi}_x & 0 & -\dot{\phi}_z \\ \dot{\phi}_x & -\dot{\phi}_y & \dot{\phi}_z & 0 \end{bmatrix},$$

and

$$E(t) = [E_1(t),\ E_2(t),\ E_3(t),\ E_4(t)]^T.$$

$\dot{\Phi}_x$, $\dot{\Phi}_y$, $\dot{\Phi}_z$ are the rates of rotation about the vehicle axes and are considered constant over each integration time interval. With this assumption the closed form solution of the differential equation is given by

$$E(t) = A(t)\ E(0) \tag{4}$$

where,

$$A(t) = \begin{bmatrix} \cos\frac{c}{2}t & -\frac{\dot{\Phi}_z}{c}\sin\frac{ct}{2} & -\frac{\dot{\Phi}_y}{c}\sin\frac{ct}{2} & -\frac{\dot{\Phi}_x}{c}\sin\frac{ct}{2} \\ \frac{\dot{\Phi}_z}{c}\sin\frac{ct}{2} & \cos\frac{ct}{2} & -\frac{\dot{\Phi}_x}{c}\sin\frac{ct}{2} & -\frac{\dot{\Phi}_y}{c}\sin\frac{ct}{2} \\ \frac{\dot{\Phi}_y}{c}\sin\frac{ct}{2} & \frac{\dot{\Phi}_x}{c}\sin\frac{ct}{2} & \cos\frac{ct}{2} & -\frac{\dot{\Phi}_z}{c}\sin\frac{ct}{2} \\ \frac{\dot{\Phi}_x}{c}\sin\frac{ct}{2} & \frac{\dot{\Phi}_y}{c}\sin\frac{ct}{2} & \frac{\dot{\Phi}_z}{c}\sin\frac{ct}{2} & \cos\frac{ct}{2} \end{bmatrix}$$

$$c^2 = \dot{\Phi}_x^{\,2} + \dot{\Phi}_x^{\,2} + \dot{\Phi}_z^{\,2} \tag{5}$$

$E(0)$ = the value of $E(t)$ at $t = 0$.

Since the exact solution is known, the results of a simulation can be compared with that of the exact solution and the difference in the two will be the error.

The errors in the numerical integration scheme vary among other factors with the order of the integration scheme employed, the integration frequency being employed, computer word length, etc. In particular the truncation errors and round-off errors vary as shown in

Figure 1.

In the truncation region the computational error is a function of both the order of the integration scheme and the integration frequency. Increased integration frequency results in a lower computational error. Increasing the order of the numerical integration scheme also reduces the computational truncation error and increases the slope of the truncation line. The slope of the truncation line in this figure is equal to the order of the employed integration scheme; e.g., for a fourth order scheme, the slope is 4; for a rectangular integration scheme, the slope is unity.*

As the integration frequency is further increased the computational error enters the region of round-off errors (neglecting quantization errors). In this region, the computational error is inversely proportional to both the computational integration step size (a slope of -1) and the computer word length (the addition of another bit to the computer word length decreases the round-off error by a factor of two).*

---

* For more details refer to [2].

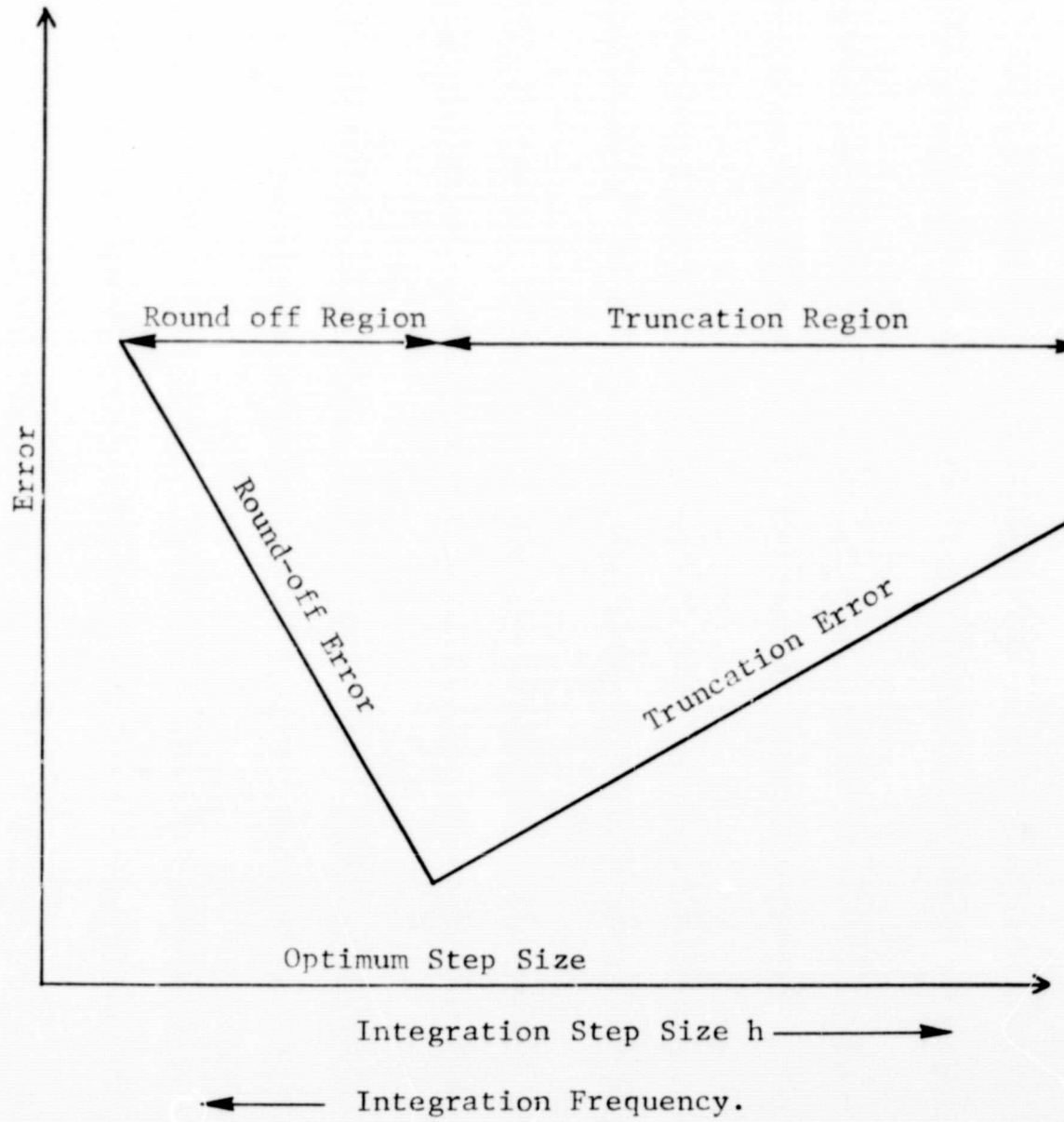


Figure 1. Regions of computational error.

III. EFFECT OF NOISE ON THE ERRORS

This chapter analyzes the effect on the error of additive random noise in the input of Equation (3). The results of simulations to determine the optimum step size for the trapezoidal integration scheme using the S/360 continuous system modeling program (S/360 C.S.M.P.) are given. The inputs for these simulations are free from any noise. These simulations are compared with simulations with noise inputs to determine the effect of noise on the errors.

The S/360 C.S.M.P. is a problem-oriented program designed to facilitate the digital simulation of continuous processes on large-scale digital machines. The program uses single precision accuracy and is very simple to implement.

In the simulation of Equation (3) the integration step size h is varied. The initial condition for the equation is

$$E(0) = [1, 0, 0, 0]^T$$

The inputs used are $\dot{\phi}_x = 2\pi t$, $\dot{\phi}_y = 2\pi t$, $\dot{\phi}_z = \pi t$. These inputs are very large compared to the specifications of the vehicle gyros which have a coarse level of control of 1°/sec and a fine level of control of .1°/sec. The large input levels were chosen to save the computer execution time.

A simulation run time of 2 secs. is used at which time the theoretical result indicates the system should return to the initial position of $[1, 0, 0, 0]^T$. Any difference between the computed value and actual

value is considered to be the error. A plot of the error vs. step size h is shown in Figure 2. It can be seen that the minimum error occurs at a step size of 8×10^{-4} secs.

Any measurement by a measuring device is not accurate and there is always a difference between the measured quantity and the actual quantity. This difference or error may be caused by a number of factors such as the accuracy of the instrument, noise introduced by the source while making the measurements, etc. These errors are random in nature and are assumed to be normally distributed with zero mean. It is therefore appropriate to investigate the effect on the errors of additive random noise in the measurements.

Next the results of simulations with gaussian noise having zero mean and a standard deviation of 3×10^{-3} degrees/sec. superimposed upon the input are given. The subroutine 'Gauss' is used to generate the noise and a sample of ten different runs is used. The average of the errors resulting from these ten samples is calculated. Table 1 lists the results of these simulations. The variance is of the order of 10^{-5} and for this variance the errors introduced by the numerical integration in component E, are predominant compared to the input noise level and hence the results of E_1 remain uneffected. On the other hand the errors in E_2, E_3 and E_4 are of the same order as the input and are effected as can be seen from Table 1.

To correlate the noise level and the errors additional simulations were made for different values of variance keeping the other variables the same. The standard deviation of the output was calculated as follows:

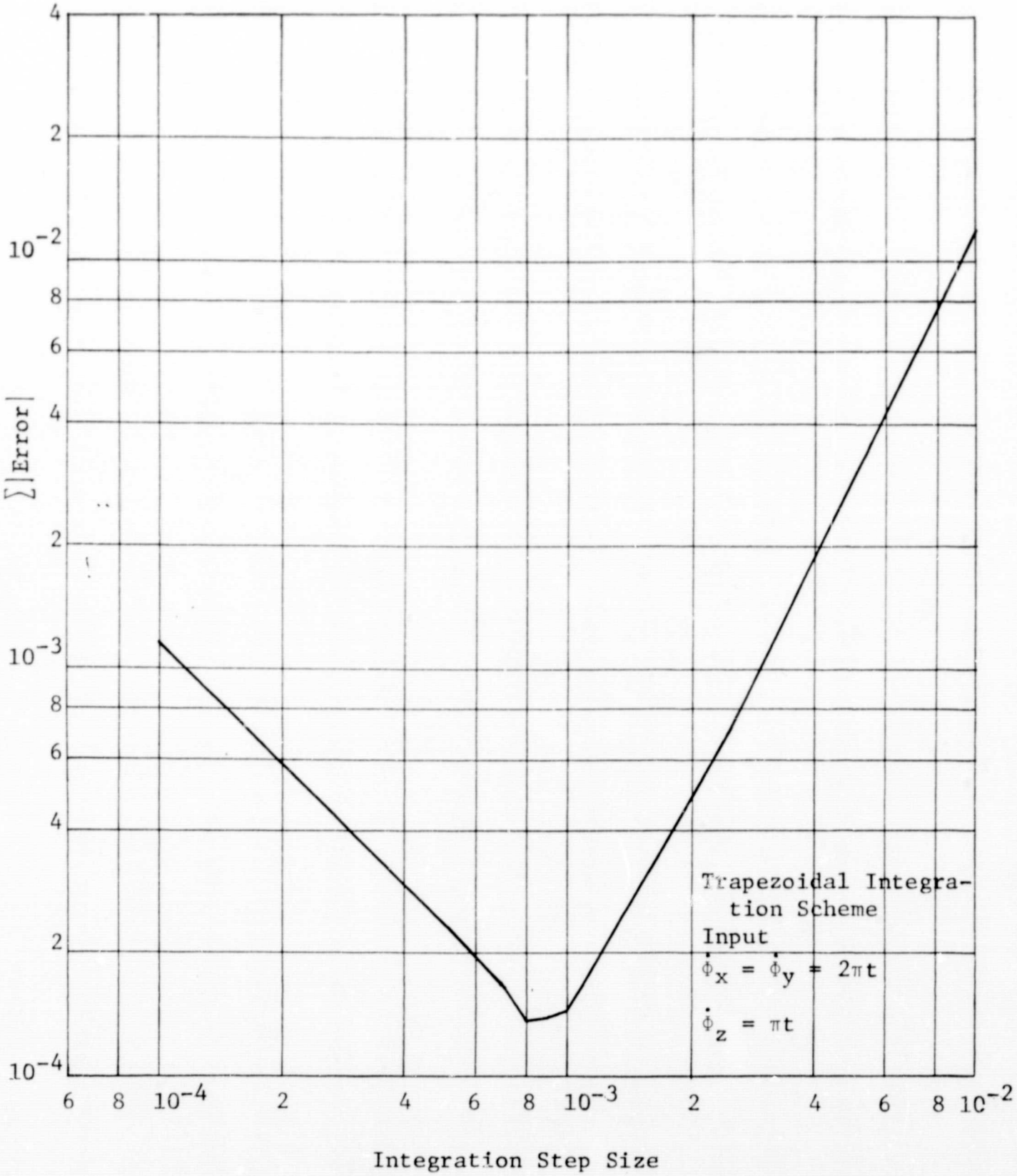


Figure 2. Error vs. integration step size using S/360 C.S.M.P.

TABLE 1

COMPARISON OF SIMULATION RESULTS WITH AND WITHOUT NOISY INPUTS

Inputs $\dot{\phi}_x = \dot{\phi}_y = 2\pi t$; $\dot{\phi}_z = \pi t$

Standard deviation of the input noise = 3×10^{-3}

	Analytical Result	Without Noisy Input	Average of Output with Noisy Input
E_1	-1.0000	-9.9989×10^{-1}	-9.9989×10^{-1}
E_2	0	1.2936×10^{-6}	1.1221×10^{-6}
E_3	0	-2.0351×10^{-5}	-1.1821×10^{-5}
E_4	0	4.6417×10^{-6}	7.201×10^{-6}

$$(\text{Standard deviation})^2 = E(x^2) - E^2(x)$$

where,

x is the random variable whose standard deviation is to be calculated.

E stands for the expected value or mean.

The resulting standard deviation is plotted against the input standard deviation in Figure 3. The analysis of these results shows that as expected the errors increase as the variance of the noise is increased.

Conclusions

From the above analysis it appears that the noisy inputs affect the computation errors if the input noise level is comparable to the computation errors without noisy inputs. Henrici [3] has suggested that the random errors be assumed as having a normal distribution. With this assumption a better estimate of the output is possible using noisy inputs and is a possible area of future research. The effect of noise on the optimum step size and the correlation of the output noise level to the input noise level are two other areas for future work.

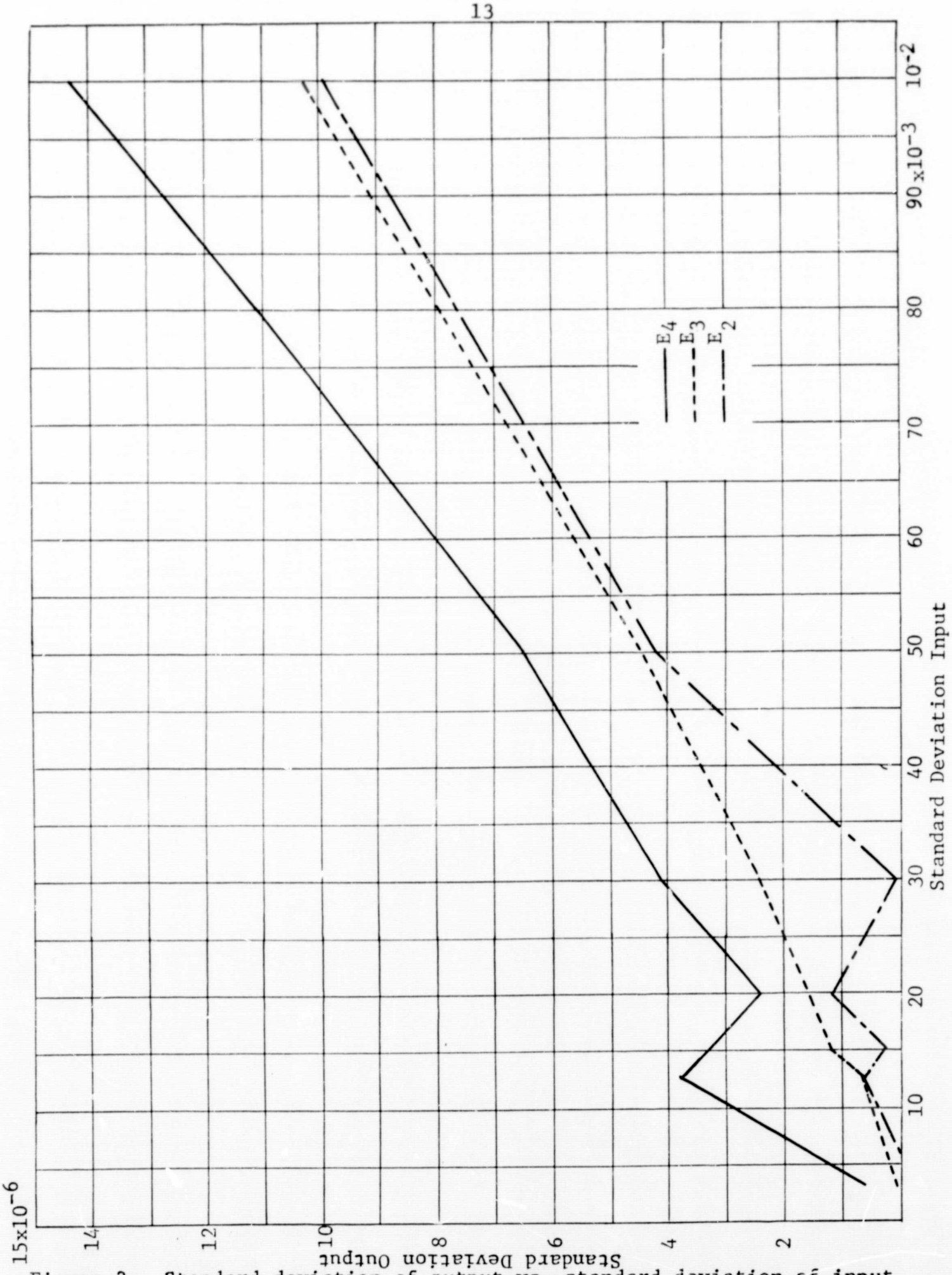


Figure 3. Standard deviation of output vs. standard deviation of input.

IV. ERRORS IN FIXED POINT ARITHMETIC OPERATION

This chapter considers the simulation of differential equations in the fixed point mode of arithmetic operations using double precision. The results obtained are compared with those of the floating point mode of operation and certain conclusions are drawn for the fixed point mode of operations. Effect of change in the input level on errors and on the optimum step size is also investigated in these simulations.

Comparison of Errors in Fixed Point and Floating Point Mode of Operation

The program simulates a fixed point machine using Fortran IV programming language on a floating point machine. The 16-bit word size double precision computer is used to integrate $\dot{E}$ to calculate the value of E.

In the fixed point mode of operation the trapezoidal integration scheme is used. This plot of error vs. integration step size indicates that the optimum step size is 7×10^{-4} secs. and that the magnitude of the total error for this step size is 2×10^{-4} as shown in Figure 4. Recalling the simulation results of Chapter III, one observes that these results are of the same order as those obtained using the floating point mode of operations (optimum step size 8×10^{-4} and total error = 1.36×10^{-4}). This is not surprising, since the truncation errors are predominant and remain the same in the floating point and fixed point mode of operations.

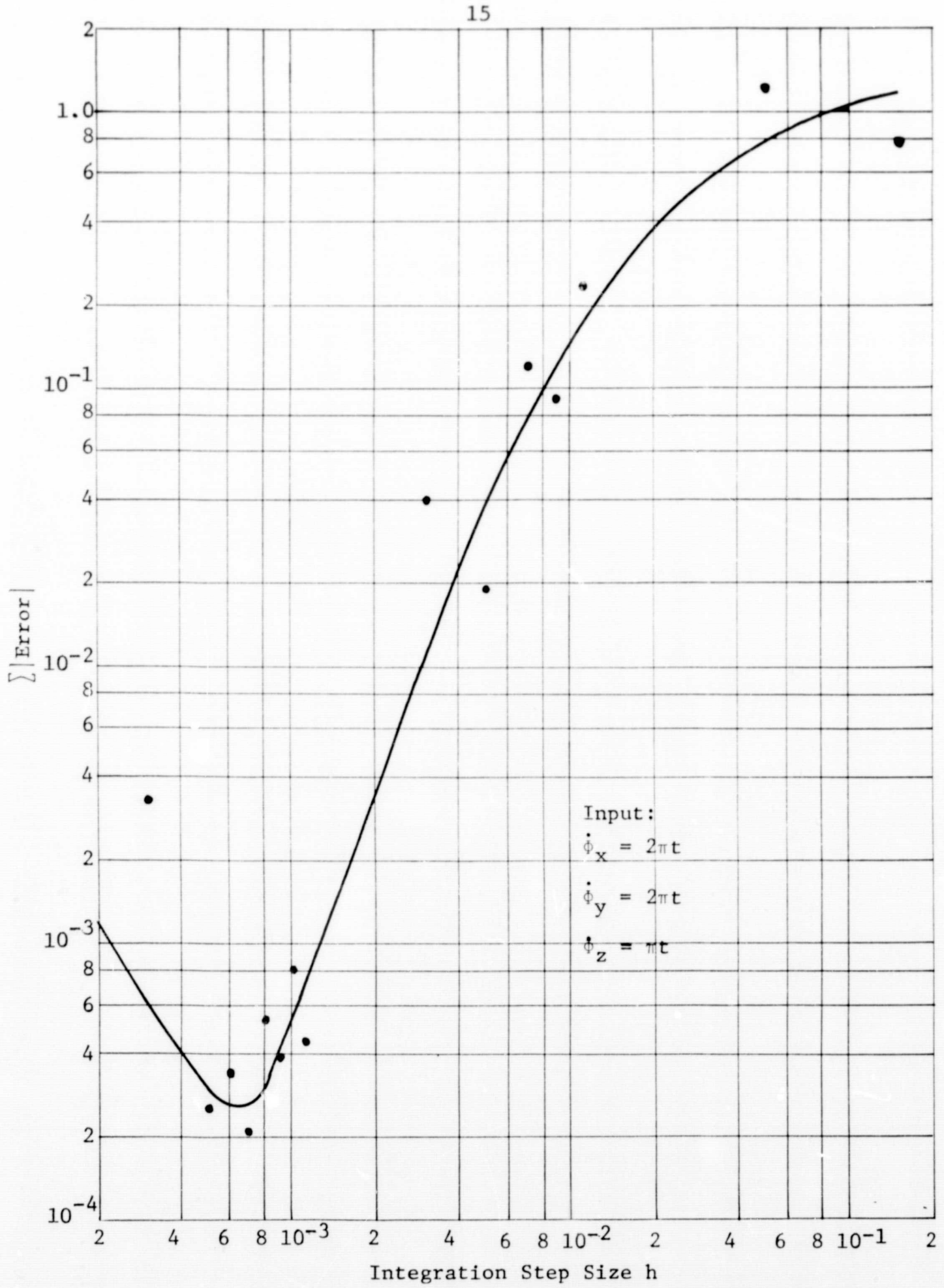


Figure 4. Error vs. integration step size using fixed point mode of operation.

Effect of Change in Input Levels

Further simulations were run in both the floating point and fixed point modes to determine the effect of change in input levels on the errors and on the optimum step size. Simulations were run in floating point mode using the Trapezoidal integration scheme with inputs of 10 times the previous value

$$\dot{\Phi}_x = 20\pi t \ , \ \dot{\Phi}_y = 20\pi t, \ \dot{\Phi}_z = 10\pi t.$$

The simulation results with these two inputs are plotted on the same graph The optimal step size shifts from 8×10^{-4} secs. to 1.5×10^{-4} secs. as shown in Figure 5. The total error (absolute sum of the elements of the error matrix) increases from 1.36×10^{-4} to 4.6×10^{-3}, an increase of 33.77. This is well within the expected range since the number of computations has increased by a factor of 5.33 while the input magnitudes have increased by a factor of 10, which yields an expected increase in error by a factor of 53.3. As discussed earlier, the change of mode of operation does not effect the results appreciably and a similar analysis in the fixed point mode should indicate similar results. This is verified when the input level is reduced from $\dot{\Phi}_x = \dot{\Phi}_y = 2\pi t$ radians/second, $\dot{\Phi}_z = \pi t$ radians/sec. to $\dot{\Phi}_x = \dot{\Phi}_y = \frac{\pi t}{180}$ radians/sec., $\dot{\Phi}_z = \frac{\pi t}{360}$ radians/sec. (equivalent to the coarse level input to which the vehicle Gyros will be subjected). The results are shown in Figure 6 and indicate that the optimum step size shifts from 7×10^{-4} to 3×10^{-2} and that the total error reduces from 2.5×10^{-4} to 5×10^{-8}, a decrease of $5 \times 10^{+3}$.

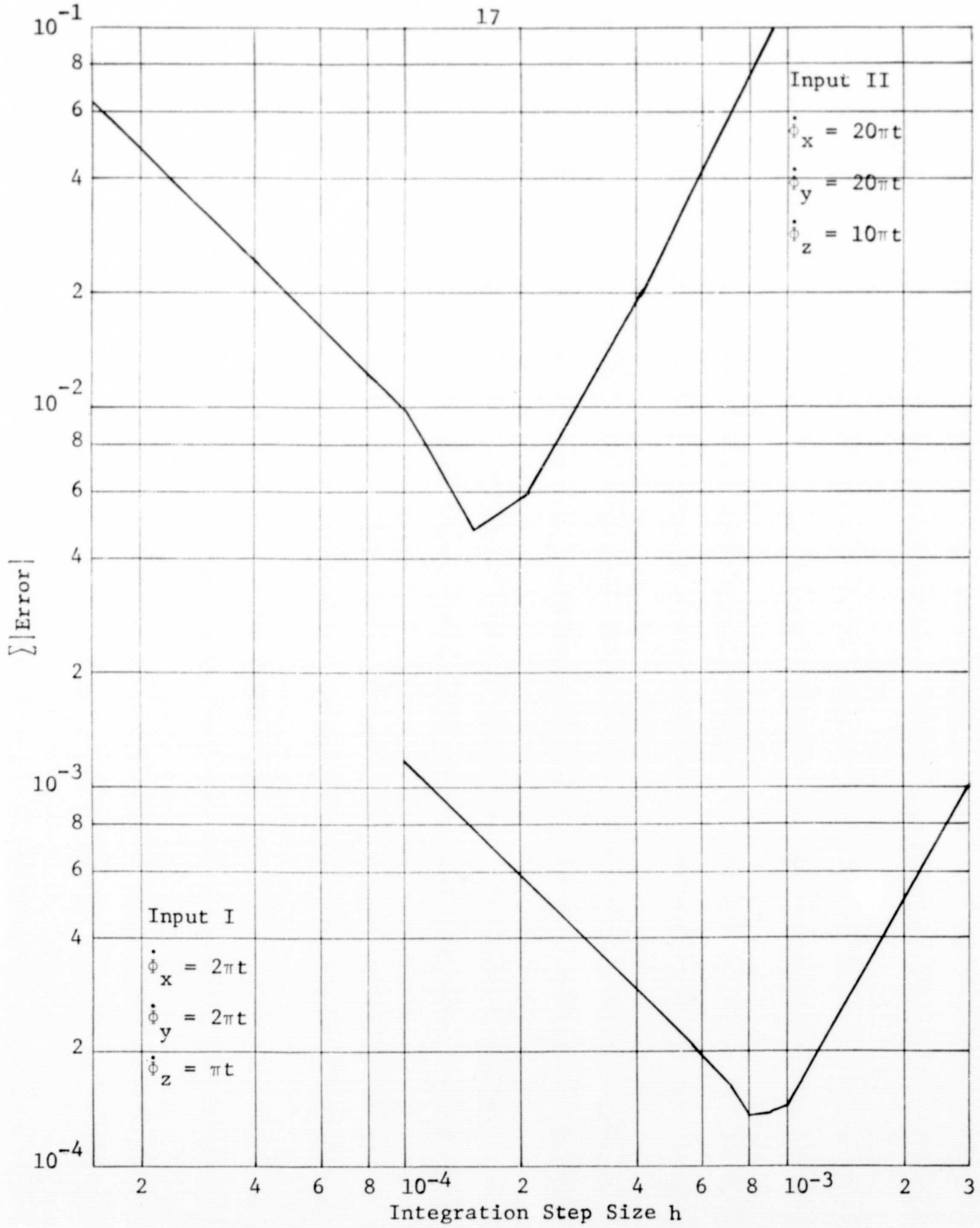


Figure 5. Error vs. h for two different inputs using floating point mode of operation.

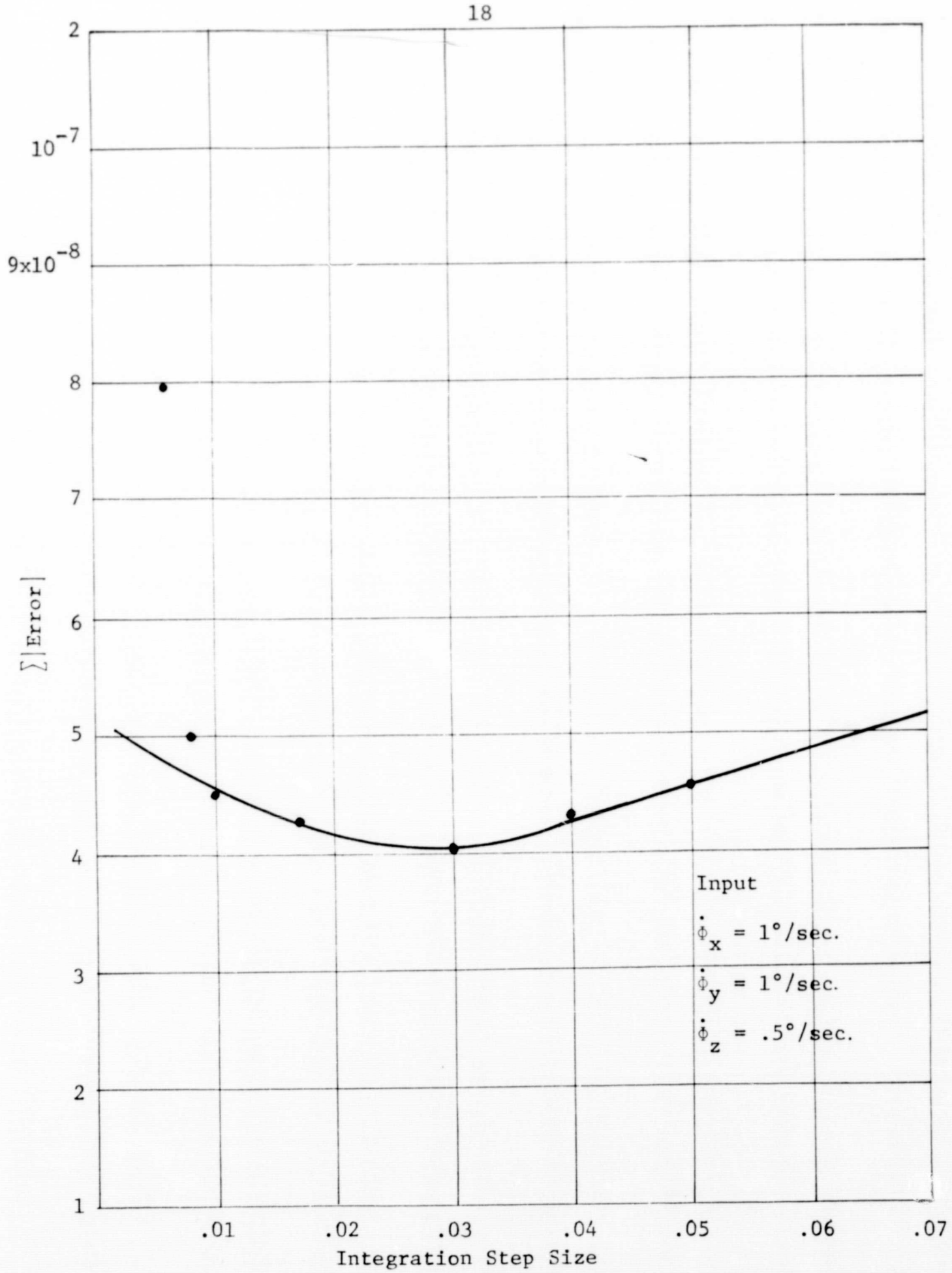


Figure 6. Error vs. integration step size for fixed point mode of operation.

Since the input has been reduced by a factor of 360 and the number of integrations increased by 40, a decrease in the error by a factor of 15×10^3 is expected. Thus, the results agree with the expected values.

Assuming that the vehicle is rotating about the x axis ($\dot{\phi}_x$ being nonzero and $\dot{\phi}_y = \dot{\phi}_z = 0$), the matrix A(t) of Equation (4) reduces to

$$A(t) = \begin{bmatrix} \text{Cos } \frac{ct}{2} & 0 & 0 & -\text{Sin } \frac{ct}{2} \\ 0 & \text{Cos } \frac{ct}{2} & -\text{Sin } \frac{ct}{2} & 0 \\ 0 & \text{Sin } \frac{ct}{2} & \text{Cos } \frac{ct}{2} & 0 \\ \text{Sin } \frac{ct}{2} & 0 & 0 & \text{Cos } \frac{ct}{2} \end{bmatrix} \tag{7}$$

If in Equation (4) the initial conditions are $E(0) = [1, 0, 0, 0]^T$, the solution of Equation (3) reduces to

$$E_1(t) = \text{Cos } \frac{ct}{2} \tag{8}$$

$$E_4(t) = \text{Sin } \frac{ct}{2} \tag{9}$$

where $c = \dot{\phi}_x$. Assuming that the vehicle is rotating about the y axis with the same initial conditions as above one obtains as the solution of (3):

$$E_1(t) = \text{Cos } \frac{ct}{2} \tag{10}$$

$$E_3(t) = \text{Sin } \frac{ct}{2} \tag{11}$$

where $c = \dot{\phi}_y$. Similarly, if the vehicle is assumed to be rotating

about the z axis with the same initial conditions, the solution of (3) reduces to

$$E_1(t) = \text{Cos } \frac{ct}{2} \tag{12}$$

$$E_2(t) = \text{Sin } \frac{ct}{2} \tag{13}$$

where

$$c = \dot{\phi}_z$$

Keeping the value of c the same in the equations (8) to (13) yields similar results and hence the errors should be the same in all cases. Simulations using fixed point mode of operations verify this conclusion and the error vs. the integration step size h is the same in all three cases as shown in Figure 7. A constant input of 1°/sec. yields an optimal step size of .12 secs. with the total error 6.5×10^{-6}.

Recalling the conclusions drawn from Figure 6 (that in the fixed point mode of operation reducing the input increases the optimal step size while reducing the errors), one expects the optimal step size to increase and errors to decrease when the input level is reduced. The simulations verify this conclusion. Reducing the input from 1°/sec. to $.1^\circ$/sec. (the fine level of control to which the vehicle gyros are sensitive) the optimal step size increases to .2 secs. and the total error reduces to 45×10^{-8}. These results are plotted in Figure 8.

When the constant input of $.1^\circ$/sec. is given to three axes at a time, using the fixed point mode of operation, one obtains the total error of 81×10^{-8} for the optimum step size of .17 secs. This result is plotted in Figure 9. The optimal step size has thus shifted from

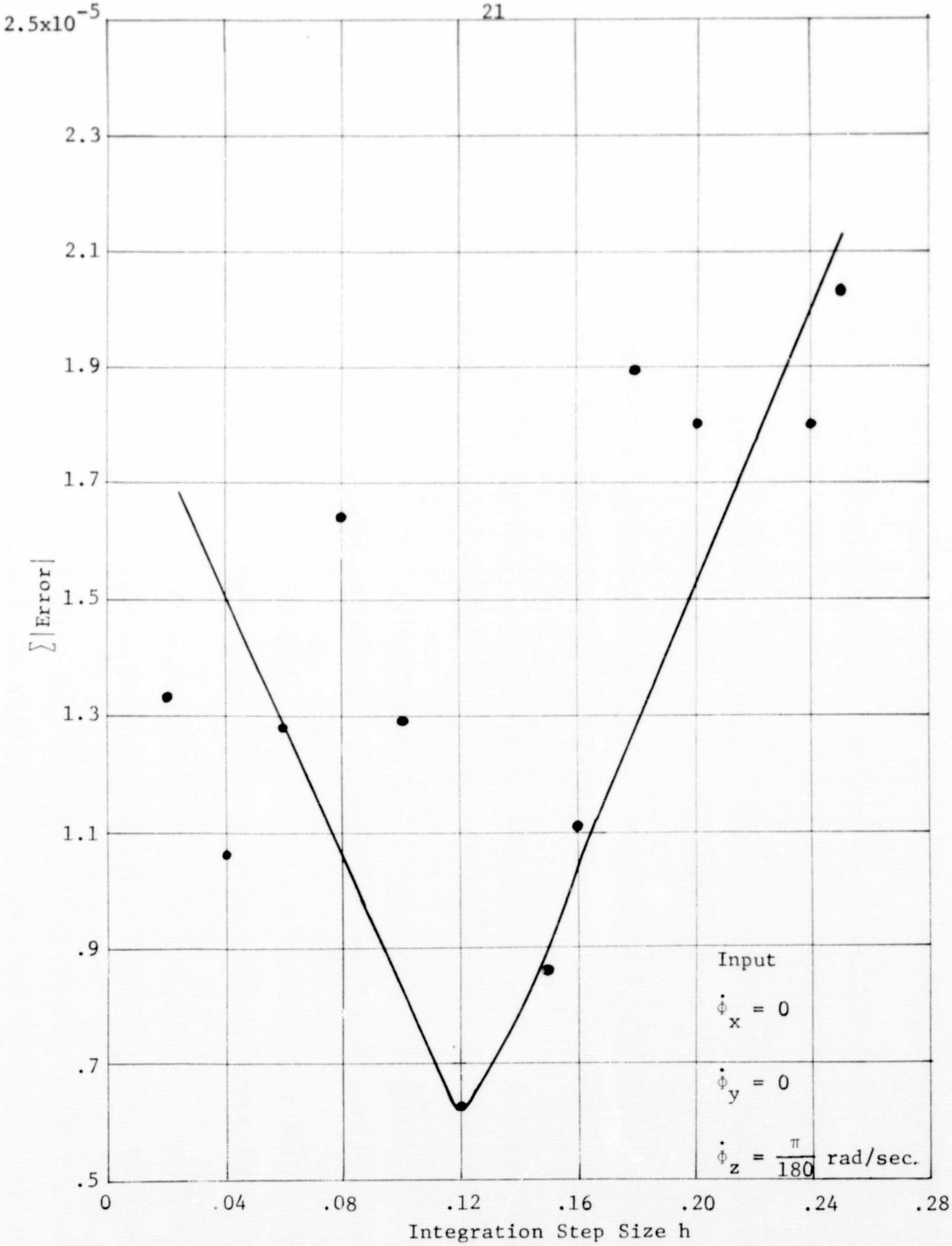


Figure.7. Error vs. integration step size for input of 1°/sec. given to one axis only.

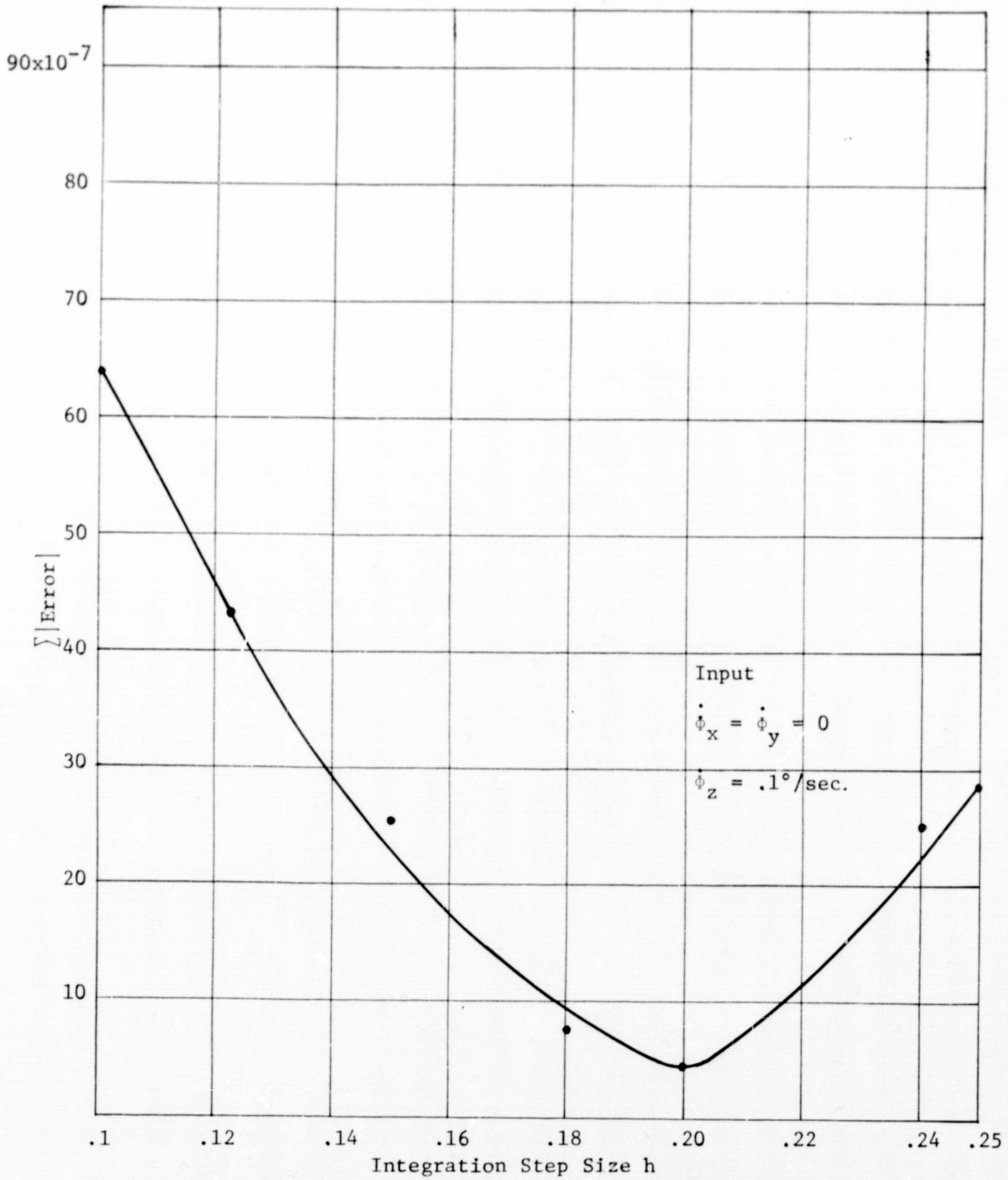


Figure 8. Error vs. h for input of .1°/sec. on one axis only.

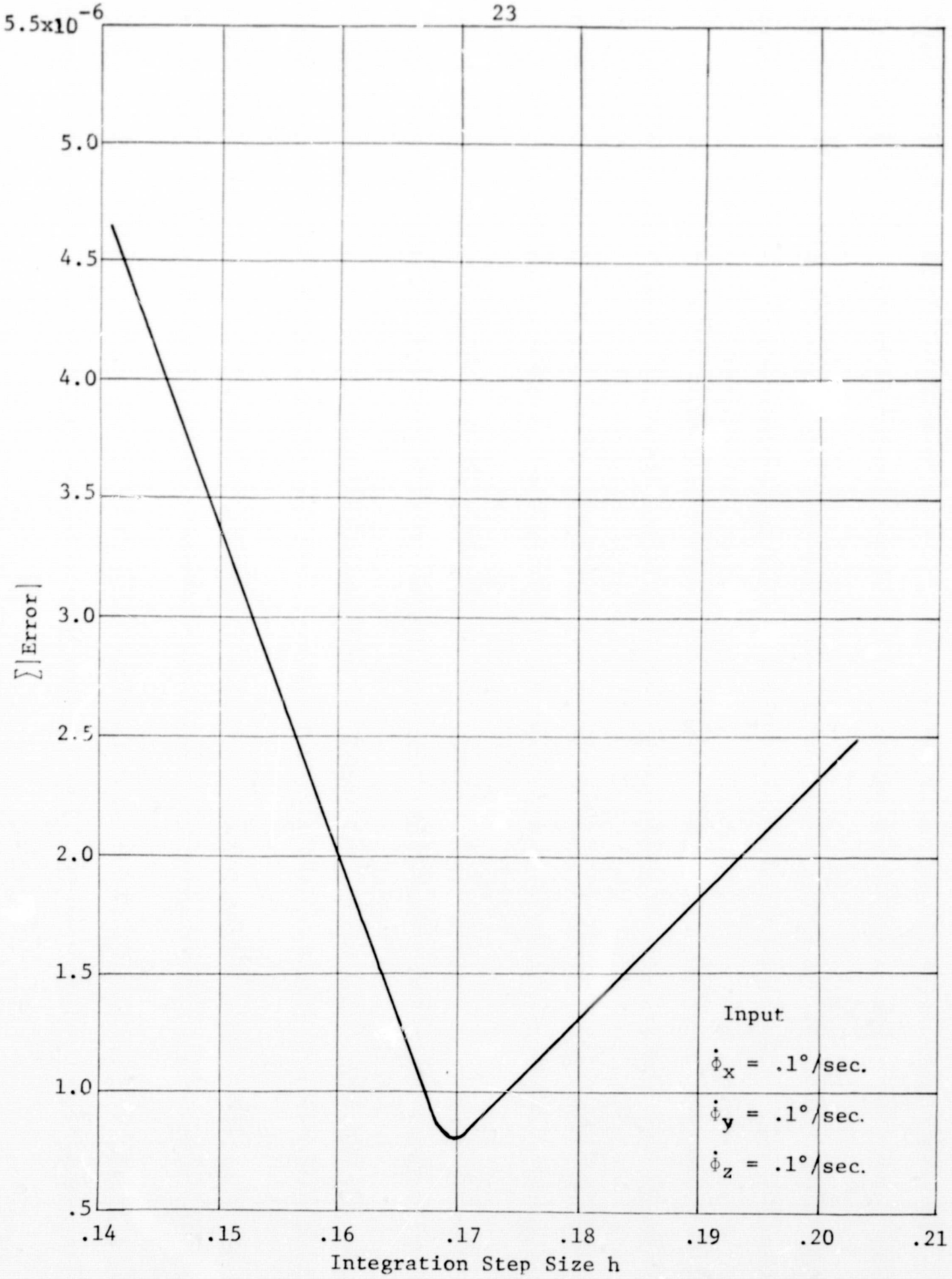


Figure 9. Error vs. integration step size when all three axis are given the same input of .1°/sec.

.2 secs. (for the same input to one axes only) to .17 secs. in the present case while the errors increase from 45×10^{-8} to 81×10^{-8}.

Conclusions

It can be concluded from the above results that the optimum step size is dependent upon the level of the input. As the input magnitude is increased the optimum step size will decrease but the magnitudes of errors at this optimum step size will increase.

V. STABILITY THEOREMS

In this chapter error bounds in the numerical solution of Equation (1) are obtained using some theorems in practical stability. After some basic definitions are presented; several theorems on practical stability are given. Proofs are presented for those theorems which do not have formal proofs in the literature. These theorems are then applied to several examples to demonstrate their effectiveness in obtaining trajectory bounds of a particular class of systems.

Definitions

Let $t_i \in J$ and let $x_i = x(t_i; x_i, t_i)$. The following definitions will be applied in the development to follow:

DEFINITION 1. System (1) is stable with respect to $\{S_o(t), S(t), t_o\}$. if $x_o \in S_o(t_o)$ implies $x(t; x_o, t_o) \in S(t)$ for all $t \in J$.

DEFINITION 2. System (1) is uniformly stable with respect to $\{S_o(t), S(t)\}$ if for each $t_i \in J$, $x_i \in S_o(t_i)$ implies $x(t; x_i, t_i) \in S(t)$ for all $t \in [t_i, \infty)$

DEFINITION 3. System (1) is practically stable with respect to $(\alpha, \beta, t_o, ||\cdot||)$, $\alpha \leq \beta$ if $||x_o|| < \alpha$ implies $||x(t; x_o; t_o)|| < \beta$ for all $t \in J$.

DEFINITION 4. System (1) is uniformly practically stable with respect to $(\alpha, \beta, ||\cdot||)$, $\alpha \leq \beta$ if for each $t_i \in J$, $||x_i|| < \alpha$ implies

$||x(t; x_i, t_i)|| < \beta$ for all $t \in [t_i, \infty)$.

In the definitions above the different time varying sets and trajectories are as shown in Figure 10.

The following definition is from Kamke [4].

DEFINITION 5. A vector function $f = (f_1, f_2, \ldots, f_n)$ of a vector variable $x = (x_1, \ldots, x_n)$ will be said to be of type K in a set S if for each subscript $i = 1, \ldots n$, $f_i(a) \leq f_i(b)$ for any two points $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$ in S with $a_i = b_i$ and $a_k \leq b_k$ $(k = 1, \ldots, n, k \neq i)$.

By the above definition it is evident that every scalar function is of type K since the condition is satisfied trivially for n = 1. However, a vector function (f_1, f_2) of two variables (x_1, x_2) is of type K iff f_1 is a non-decreasing function of x_2 and f_2 is a non-decreasing function of x_1.

Theorems on Practical Stability

First, a preliminary result is stated in the form of Theorem 1 which is taken from Coppel [5].

THEOREM 1. Assume for the scalar differential equation

$$\dot{y} = g(y, t) \tag{14}$$

$g(y, t)$ is defined for all $t \in J$. Further let $g(y, t)$ be of type K for each fixed value of t. Let $y(t; y_o, t_o)$ with $y_o = y(t_o; y_o, t_o)$ denote a solution of (14) which exists on a closed interval $[t_o, t_a]$.

Suppose there exist two scalar functions q(t), and r(t) continuous on $[t_o, t_a]$ which satisfy the inequalities

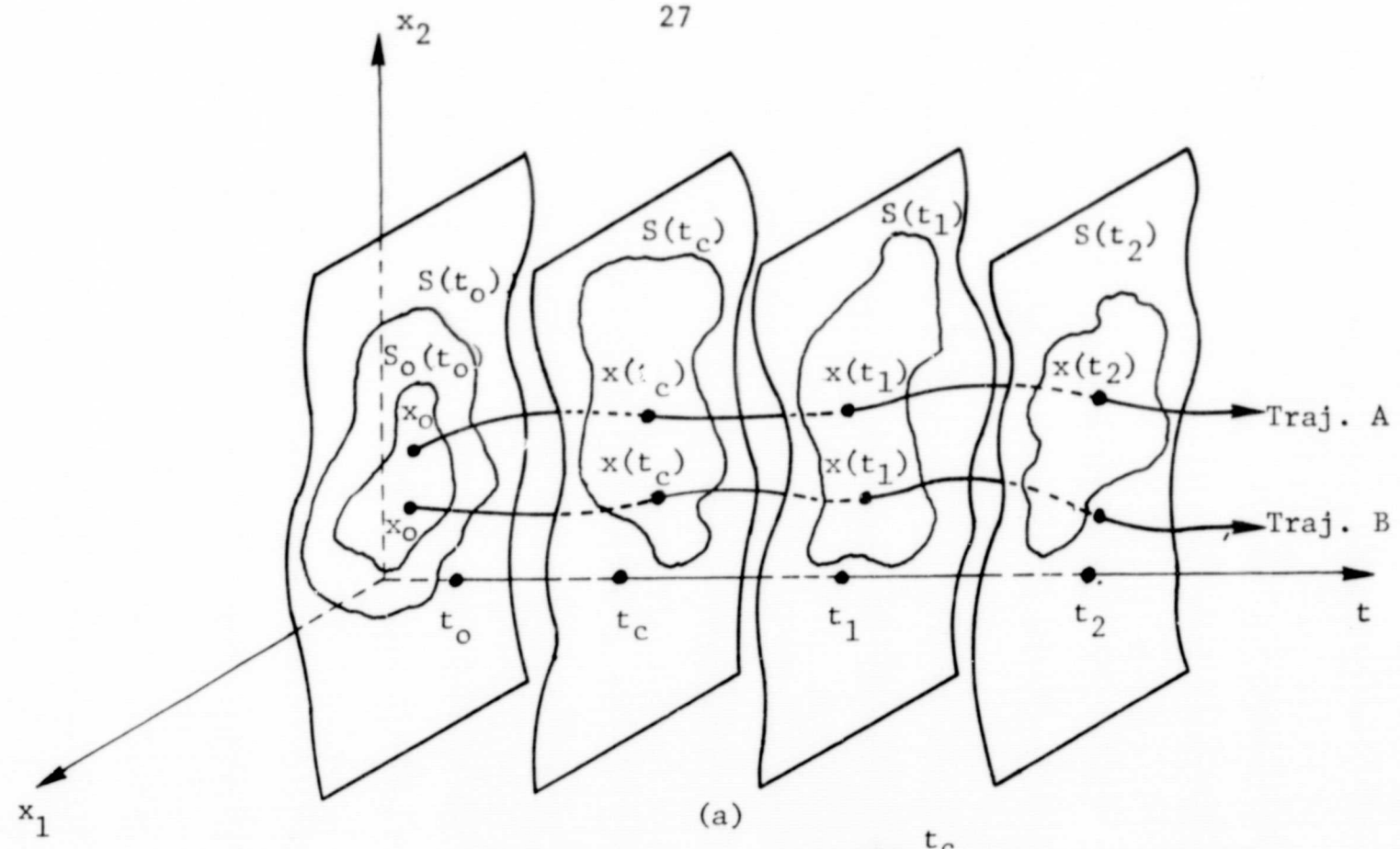


Figure 10. Different trajectories for a second order system.

(a) Trajectory A stable with respect to $\{S_o(t_o), S(t), t_o\}$
Trajectory B unstable with respect to $\{S_o(t_o), S(t), t_o\}$

(b) Practically stable with respect to $\{\alpha, \beta, t_o, ||\cdot||\}$

(c) Practically unstable with respect to $\{\alpha, \beta, ||\cdot||\}$.
Asymptotically stable in Liapunov sense (Qualitative) and practically unstable in the sense of quantative definition.

$$\dot{q}(t) > g(q(t), t)$$

$$\dot{r}(t) < g(r(t), t).$$

on the half open interval $(t_o, t_a]$ such that

$$r(t_o) \leq y_o \text{ and } q(t_o) \geq y_o.$$

Then,

$$r(t) < y(t; y_o, t_o)$$

and

$$q(t) > y(t; y_o, t_o) \text{ for all } t \in (t_o, t_a]$$

or

$$r(t) < y(t; y_o, t_o) < q(t) \text{ for } t \in (t_o, t_a].$$

PROOF. By continuity $q(t) > y(t; y_o, t_o)$ on an interval $[t_o, t_o + \delta]$ where $\delta > 0$. Assume that the inequality

$$q(t) > y(t; y_o, t_o)$$

does not hold throughout the interval $[t_o, t_a]$, then there would exist a value $t_c (t_o \leq t_c \leq t_a)$ such that

$$q(t) > y(t; y_o, t_o) \text{ for } t_o \leq t < t_c \text{ and}$$

$$q(t_c) \geq y(t_c)$$

and

$$q_i(t_c) \geq y_i(t_c) \tag{15}$$

for at least one i. But

$$\dot{q}_i(t_c) > g_i(q_i(t_c), t_c) \qquad \text{(given in theorem)}$$

$$> g_i(y_i(t_c), t_c) \text{ therefore}$$

$$\dot{q}_i(t_c) > \dot{y}_i(t_c). \tag{16}$$

Since $q_i(t_c) = y_i(t_c)$ for at least one i it follows form (15) that

$$q_i(t) < y_i(t) \tag{17}$$

for a value of t less than but arbitrarily close to t_c. But this contradicts the definition of t_c in the assumption. Therefore

$$q(t) > y(t;\ y_o,\ t_o)$$

Similarly it can be proved that

$$r(t) < y(t;\ y_o,\ t_o)$$

by a change of variables of y to -y and q to - q.

The following theorems are taken from Michel [6] and stated here without proof.

THEOREM 2. System (1) is stable with respect to $\{S_o(t),\ S(t),\ t_o\}$ $S(t) \supset S_o(t)$ and $\delta S(t) \cap \delta S_o(t) = \phi$ for $t \in J$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in S(t)$, for all $t \in J$, and a function g which satisfies the assumptions of Theorem 1, such that

(i) $\dot{V}_{(1)}(x, t) < g(V, t) \quad x \in S(t),\ t \in J$

(ii) $y(t;\ \sup_{x \in S_o(t_o)} V(x, t_o),\ t_o) \leq \inf_{x \in \delta S(t)} V(x, t),\ t \in J$

($y(t;\ y_o,\ t_o)$ is as given in Theorem 1.)

THEOREM 3. System (1) is uniformly stable with respect to $\{S_o(t),\ S(t)\}$,

$$S(t) \supset S_o(t) \text{ and } \delta S(t) \cap \delta S_o(t) = \phi$$

for all $t \in J$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in [S(t) - \overline{S_o(t)}]$, for all $t \in J$, and a function g which satisfies the assumption of Theorem 1, such that

(i) $\dot{V}_{(1)}(x, t) < g(V, t) \qquad x \in [S(t) - \overline{S_o(t)}],\ t \in J.$

(ii) $y(t_2; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) \leq \inf_{x\varepsilon\delta S(t_2)} V(x, t_2); \; t_1, t_2 \varepsilon J,$

$t_2 > t_1$

(y is as defined in Theorem 1.)

THEOREM 4. System (1) is stable with respect to $\{S_o(t), S(t), t_o\}$ $S(t_o) \supset S_o(t)$ for all $t \varepsilon J$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \varepsilon S(t)$, for all $t \varepsilon J$, and a real valued function $\psi(t)$ which is integrable over J, such that

(i) $\dot{V}_{(1)}(x, t) < \psi(t), \; x \varepsilon S(t), \; t \varepsilon J.$

(ii) $\int_{t_o}^{t} \psi(\tau)d\tau \leq \inf_{x\varepsilon\delta S(t)} V(x, t) - \sup_{x\varepsilon S_o(t_o)} V(x, t_o), \; t \varepsilon J.$

The following Theorems are stated without proof in form of Theorems and corollaries in Michel [6]. The proofs of these have been fully developed here.

THEOREM 5. System (1) is uniformly stable with respect to $\{S_o(t), S(t)\}$ $S(t) \supset S_o(t)$ and $\delta S(t) \cap \delta S_o(t) = \phi$ for all $t \varepsilon J$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \varepsilon [S(t) - \overline{S_o(t)}]$, for all $t \varepsilon J$, and a real valued function $\psi(t)$ which is integrable over J, such that

(i) $\dot{V}_{(1)}(x, t) < \psi(t), \; x \varepsilon [S(t) - \overline{S_o(t)}], \; t \varepsilon J$

(ii) $\int_{t_1}^{t_2} \psi(\tau)d\tau \leq \inf_{x\varepsilon\delta S(t_2)} V(x, t_2) - \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1),$

$t_1, t_2 \varepsilon J; \; t_2 > t_1.$

PROOF. In Theorem 1 let $g(y, t) \equiv \psi(t)$ for all $t \varepsilon J$ so that

$$\dot{y} = \psi(t) \text{ and } \dot{r}(t) < g(r(t), t)$$

$$< \psi(t)$$

Let,

$$r(t) = V[x(t; x_o, t_o), t] \text{ so that}$$

$$\dot{V}_{(1)}\,[x(t; x_o, t_o), t] < g(V, t)$$

$$\dot{V}_{(1)}\,[x(t; x_o, t_o), t] < \psi(t).$$

If $\psi(t)$ is integrable over J, then,

$$y(t_2; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1)$$

$$= y(t_1; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) + \int_{t_1}^{t_2} \psi(\tau)d\tau \qquad (18)$$

or

$$\int_{t_1}^{t_2} \psi(\tau)d\tau$$

$$= y(t_2; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) - y(t_1; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) \qquad (19)$$

From Theorem 1

$$y(t_1; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) \geq \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1).$$

Considering hypothesis (ii) of Theorem 3

$$y(t_2; \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1), t_1) \leq \inf_{x\varepsilon\delta S(t_2)} V(x, t_2);\ t_1, t_2 \varepsilon J \text{ and}$$

$$t_2 > t_1$$

Therefore

$$\int_{t_1}^{t_2} \psi(\tau)d\tau \leq \inf_{x\varepsilon\delta S(t_2)} V(x, t_2) - \sup_{x\varepsilon\delta S_o(t_1)} V(x, t_1) \qquad (20)$$

which proves the theorem.

THEOREM 6. System (1) is practically stable with respect to $(\alpha, \beta, t_o, ||\cdot||)$, $\alpha < \beta$, if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in B(\beta)$, for all $t \in J$, and a real valued function $\psi(t)$ which is integrable over J, such that

(i) $\dot{V}_{(1)}(x, t) < \psi(t)$; $x \in B(\beta)$, $t \in J$

(ii) $\int_{t_o}^{t} \psi(\tau)d\tau \leq \inf_{||x||=\beta} V(x, t) - \sup_{||x||<\alpha} V(x, t_o)$, $t \in J$.

PROOF. In Theorem 4 let

$S(t) = B(\beta)$,

$S_o(t_o) = S_o(t) = B(\alpha)$ and

$S_o(t_o) \equiv ||x|| < \alpha$, where $\alpha < \beta$.

Let,

$g(y, t) \equiv \psi(t)$ in Theorem 1 for all $t \in J$.

$\dot{y} = \psi(t)$ and $\dot{r}(t) < g(r(t), t) < \psi(t)$

Let,

$r(t) = v[x(t; x_o, t_o), t]$ so that

$\dot{r}(t) = \dot{V}_{(1)}(x, t) < g(V, t) = \psi(t)$. $x \in B(\beta)$, $t \in J$.

If $\psi(t)$ is integrable over J, one can write

$$y(t); \sup_{||x||<\alpha} V(x, t_o), t_o) = y(t_o; \sup_{||x||<\alpha} V(x,t_o), t_o) + \int_{t_o}^{t} \psi(\tau)d\tau. \quad (21)$$

or

$$\int_{t_o}^{t} \psi(\tau)d\tau = y(t; \sup_{||x||<\alpha} V(x, t_o), t_o) - y(t_o; \sup_{||x||<\alpha} V(x, t_o), t_o) \quad (22)$$

From Theorem 1,

$r(t_o) \leq y_o$.

In this case,

$$y_o = y(t_o; \sup_{||x||<\alpha} V(x, t_o), t_o)$$

and

$$r(t_o) = V[x(t_o; x_o, t_o)\, t_o] = \sup_{||x||<\alpha} V(x, t_o)$$

therefore

$$y(t_o; \sup_{||x||<\alpha} V(x, t_o), t_o) \geq \sup_{||x||<\alpha} V(x, t_o).$$

From hypothesis (ii) of Theorem 2 after substituting proper values of the sets $S_o(t_o)$ and $\delta S(t)$ one obtains

$$y(t; \sup_{||x||<\alpha} V(x, t_o), t_o) \leq \inf_{||x||=\beta} V(x, t)$$

Thus substituting these values one obtains:

$$\int_{t_o}^{t} \psi(\tau)d\tau \leq \inf_{||x||=\beta} V(x, t) - \sup_{||x||<\alpha} V(x, t_o). \tag{23}$$

which proves the theorem.

THEOREM 7. System (1) is practically stable with respect to $(\alpha, \beta, t_o, ||\cdot||)$, $\alpha < \beta$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in B(\beta)$, for all $t \in J$, such that

(i) $\dot{V}_{(1)}(x, t) < 0$. $\quad x \in B(\beta)$, $t \in J$.

(ii) $\sup_{||x||<\alpha} V(x, t_o) \leq \inf_{||x||=\beta} V(x, t)$. $\quad t_o < t$, $t \in J$.

PROOF. In the proof of Theorem 6 choose the function $\psi(t) \equiv 0$. Then,

$$\dot{y}(t) = 0 \text{ and } \dot{r}(t) < 0$$

Let,

$$r(t) = V(x, t)$$

Therefore,

$$V_{(1)}(x, t) < 0. \qquad x \in B(\beta),\ t \in J.$$

From Theorem 2,

$$y(t;\ \sup_{||x||<\alpha} V(x, t_o),\ t_o) \leq \inf_{||x||=\beta} V(x, t).$$

and from Theorem 1,

$$r(t_o) \leq y_o$$

In this case

$$r(t_o) = V(x(t_o;x_o, t_o), t_o) \leq \sup_{||x||<\alpha} V(x, t_o). \tag{24}$$

$$y_o = y(t_o;\ \sup_{||x||<\alpha} V(x, t_o),\ t_o) \tag{25}$$

Therefore,

$$\sup_{||x||<\alpha} V(x, t_o) \leq y_o$$

Since the hypothesis of Theorem 1 is satisfied one obtains,

$r(t) < y(t)$. For all $t \in J$.

or,

$$y(t;\ \sup_{||x||<\alpha} V(x, t_o),\ t_o) > r(t) = V(x, t) \geq \sup_{||x||<\alpha} V(x, t_o) \tag{26}$$

But from Theorem 2,

$$y(t;\ \sup_{||x||<\alpha} V(x, t_o),\ t_o) \leq \inf_{||x||=\beta} V(x, t) \tag{27}$$

$$\sup_{||x||<\alpha} V(x, t_o) \leq \inf_{||x||=\beta} V(x, t). \qquad t_o < t,\ t \in J. \tag{28}$$

which proves the theorem.

THEOREM 8. System (1) is uniformly practically stable with respect to $(\alpha, \beta, ||\cdot||)$, $\alpha < \beta$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in [B(\beta) - \overline{B(\alpha)}]$, for all $t \in J$, and a real

valued function $\psi(t)$ which is integrable over J such that

$$\text{(i)}\ \dot{V}_{(1)}(x, t) < \psi(t) \qquad x \in [B(\beta) - \overline{B(\alpha)}] \text{ and } t \in J$$

$$\text{(ii)}\ \int_{t_1}^{t_2} \psi(\tau)d\tau \leq \inf_{||x||=\beta} V(x, t_2) - \sup_{||x||=\alpha} V(x, t_1).$$

$$t_1, t_2 \in J \quad t_2 < t_1$$

PROOF. The proof of this theorem is exactly the same as that of Theorem 5 if the sets $S(t)$, $S_o(t)$, $\delta S_o(t_1)$ and $\delta S(t_2)$ are defined as follows:

$$S(t) \equiv B(\beta);\ S_o(t) \equiv B(\alpha)$$

$$\delta S(t_2) \equiv \{||x|| = \beta\} \text{ and } \delta S_o(t_1) \equiv \{|| x || = \alpha\}$$

THEOREM 9. System (1) is uniformly practically stable with respect to $(\alpha, \beta, ||\cdot||)$, $\alpha < \beta$ if there exists a real valued function $V(x, t)$ which is in C^1 for all $x \in [B(\beta) - \overline{B(\alpha)}]$, for all $t \in J$, such that

$$\text{(i)}\ \dot{V}_{(1)}(x, t) < 0 \qquad x \in [B(\beta) - \overline{B(\alpha)}].\ t \in J.$$

$$\text{(ii)}\ \sup_{||x||=\alpha} V(x, t_1) \leq \inf_{||x||=\beta} V(x, t_2),\ x \in [B(\beta) - \overline{B(\alpha)}]$$

$$t_2, t_1 \in J \text{ and } t_2 > t_1$$

PROOF. In the proof of Theorem 5 take the sets

$$\delta S(t_2) \text{ as } ||x|| = \beta \text{ and } \delta S_o(t_1) \text{ as } ||x|| = \alpha$$

Let

$$g(y, t) \equiv \psi(t) = 0 \text{ for all } t \in J \text{ so that}$$

$$\dot{y} = 0 \text{ and } \dot{r}(t) < g < 0.$$

Therefore from Theorem 1, $y(t) > r(t)$. Let

$$r(t) = V[x(t; x_o, t_o), t]$$

so that

$$\dot{V} < 0 \text{ for } x \in [B(\beta) - \overline{B(\alpha)}]$$

and,

$$y(t) > r(t) > V(x, t) \text{ for all } t \in J.$$

Therefore,

$$y(t_1; \sup_{||x||=\alpha} V(x, t_1), t_1) \geq \sup_{||x||=\alpha} V(x, t_1) \text{ for all } t \in J$$

From hypothesis (ii) of Theorem 2,

$$y(t; \sup_{x \in S_o(t_o)} V(x, t_o), t_o) \leq \inf_{x \in \delta S(t)} V(x, t), \; t \in J. \tag{29}$$

$$y(t_2, \sup_{x \in S_o(t_o)} V(x, t_o), t_o) \leq \inf_{||x||=\beta} V(x, t_2), \; t_2 \in J, \; t_2 > t_o. \tag{30}$$

Since g(y, t) is of type K,

$$y(t_1) \leq y(t_2) \text{ for } t_2 \geq t_1$$

Therefore,

$$\sup_{||x||=\alpha} V(x, t_1) < y(t_1; \sup_{||x||=\alpha} V(x, t_1), t_1) \leq y(t_2; \sup_{x \in S_o(t_o)} V(x, t_o), t_o)$$

$$\leq \inf_{||x||=\beta} V(x, t_2). \tag{31}$$

Therefore,

$$\sup_{||x||=\alpha} V(x, t_1) \leq \inf_{||x||=\beta} V(x, t_2); \; t_1, t_2 \in J \text{ and } t_2 > t_1 \tag{32}$$

Thus the theorem is proved.

Examples

Example 1. Consider the time invariant linear system

$$\dot{\underline{x}} = A\underline{x} + b\underline{u} \qquad y = \underline{c}^T\underline{x}$$

where

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{1}{4} & -1 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \qquad \underline{c}^T = [1,\ 0]$$

The transfer function of this system is

$$G(s) = \underline{c}^T(sI - A)^{-1} \quad \underline{b} = \frac{2}{(S + \frac{1}{2})^2}$$

Although this system is Liapunov stable, Dorato [7] has shown it to be short time unstable (because the system solution grows beyond the specified bounds).

The practical stability of this system will be considered here. To form a proper V function the following equation is solved for P.

$$A^TP + PA = - C = - I.$$

$$\begin{bmatrix} 0 & -\frac{1}{4} \\ 1 & -1 \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} + \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -\frac{1}{4} & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

Solving this equation, one obtains,

$$P_{21} = P_{12} = 2$$

and

$$P_{11} = \frac{21}{8},\ P_{22} = \frac{5}{2}\ .$$

Therefore,

$$V = \underline{x}^T P\underline{x} = \underline{x}^T \begin{bmatrix} \frac{21}{8} & 2 \\ 2 & \frac{5}{2} \end{bmatrix} \underline{x}$$

The eigenvalues of P are found by solving

$$\det[\lambda I - P] = 0$$

$$\det \begin{bmatrix} \lambda - \frac{21}{8} & -2 \\ -2 & \lambda - \frac{5}{2} \end{bmatrix} = 0$$

$$\lambda^2 - \frac{41}{8}\lambda + \frac{41}{16} = 0.$$

$$\lambda = \frac{41}{16} \pm \sqrt{(\frac{41}{16})^2 - \frac{41}{16}}$$

$$= \frac{41}{16} \pm \frac{\sqrt{26.4 - 10.25}}{2}$$

$$= 2.56 \pm 2.01$$

$$= 4.57 \text{ and } .55.$$

Hence,

$$\lambda\text{min.} = .55 \quad \lambda_{max} = 4.57$$

Now,

$$V = \frac{21}{8} x_1^2 + 4x_1x_2 + \frac{5}{2} x_2^2.$$

Therefore,

$$\dot{V}_{(1)} = \frac{21}{4} x_1x_2 + 4x_2^2 + 4x_1\dot{x}_2 + 5x_2\dot{x}_2$$

$$= \frac{21}{4} x_1x_2 + 4x_2^2 + 4x_1[-\frac{1}{4}x_1 - x_2 + 2u] + 5x_2[-\frac{1}{4}x_1 - x_2 + 2u].$$

With zero input $\dot{V}_{(1)}$ reduces to

$$\dot{V}_{(1)} = \frac{21}{4} x_1x_2 + 4x_2^2 - x_1^2 - 4x_1x_2 - \frac{5}{4} x_1x_2 - 5x_2^2$$

$$= -(x_1^2 + x_2^2).$$

Since $\dot{V}_{(1)} < 0$, condition (1) of Theorem 7 is satisfied. According to condition (ii) of Theorem 7,

$$\sup_{||x||<\alpha} V(x, t_o) \leq \inf_{||x||=\beta} V(x, t) \qquad t_o < t, \; t \in J.$$

But,

$$\sup_{||x||<\alpha} V(x, t_o) = \lambda_{max}||x||^2 \leq \lambda_{max}\alpha^2$$

and,

$$\inf_{||x||=\beta} V(x, t) = \lambda_{min}\, \beta^2$$

Therefore,

$$\lambda_{max}\alpha^2 \leq \lambda_{min}\, \beta^2$$

Assume the sets $||x|| = \beta$ and $||x|| < \alpha$ to be $||x|| = 5$ and $||x|| < 1$ as in Example 3 of Dorato [7]. On substitution one obtains,

$$4.57x1^2 \leq .55x5^2$$

$$4.57 \leq 13.75$$

Thus condition (ii) of Theorem 7 is also satisfied. Therefore, the system under consideration is practically stable with respect to $(1, 5, 0, ||\cdot||)$

Example 2. Time varying case. This example is taken from Coppel [5] and was first given by Cesari [8]. The equation under consideration is

$$\ddot{x} + \frac{2\dot{x}}{t} + x = 0.$$

This has the fundamental system of solutions $t^{-1} \sin t$, $t^{-1} \cos t$ and is uniformly and asymptotically stable (refer to Cesari [8]). The practical stability of this system will be investigated here. The given equation can be written as

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -\frac{2}{t} x_2 - x_1$$

or

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -\frac{2}{t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Thus,

$$A(t) = \begin{bmatrix} 0 & 1 \\ -1 & -\frac{2}{t} \end{bmatrix}$$

$$A^T(t) = \begin{bmatrix} 0 & -1 \\ 1 & -\frac{2}{t} \end{bmatrix}$$

$$C(t) = \frac{1}{2}\,[A(t) + A^T(t)] \qquad (*)$$

$$= \frac{1}{2}\begin{bmatrix} 0 & 0 \\ 0 & -\frac{4}{t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{2}{t} \end{bmatrix}.$$

Let $\Lambda(t)$ be the max instantaneous eigenvalue of $C(t)$. $\Lambda(t) = 0$ and $\lambda_{min} = -\frac{2}{t}$. Let the V function be

$$V(x, t) = V(x) = \ell n\, x^T x.$$

Norm is defined as $||x||^2 = x^T x$.

$$\dot{V}_{(1)}(x, t) = \dot{V}_{(1)}(x) = \frac{1}{x_1^2 + x_2^2}\,(2x_1x_2 + 2x_2\dot{x}_2)$$

$$= \frac{1}{x_1^2 + x_2^2}\,[2x_1x_2 + 2x_2(-\tfrac{2}{t}x_2 - x_1)]$$

$$= -\frac{4x_2^2}{t(x_1^2 + x_2^2)}.$$

Thus,

$$\dot{V}_{(1)}(x, t) < 0 \quad \forall\ t \in J.$$

Therefore, hypothesis (i) of Theorem 6 is satisfied. For the given V function

$$\sup_{||x||<\alpha} V(x, t_o) = \sup_{||x||<\alpha} \ell n||x|| \leq \ell n\, \alpha.$$

$$\inf_{||x||=\beta} V(x, t) = \inf_{||x||=\beta} \ell n||x|| = \ell n\beta$$

*Refer to Example 1 in Michel [6].

On application of hypothesis (ii) of Theorem 6 one obtains

$$\int_{t_o}^{t} o.d\tau \leq \ell n\beta - \ell n\alpha$$

$$\ell n\alpha < \ell n\beta \qquad \alpha < \beta.$$

The Theorem 6 is thus satisfied and hence this system is practically stable.

Example 3. Consider a system with a nonlinearity described by

$$\dot{\underline{x}} = A\underline{x} + \underline{b}f(\sigma).$$

$$\sigma = \underline{c}^T\underline{x}$$

The nonlinearity is given in a sector $K_1 \leq K \leq K_2$. The conditions for this system to be uniformly practically stable shall be investigated. From norm inequalities:

$$\sigma \leq ||c||\ ||x||.$$

Let the V function for the above system be $V(x, t) = V(x) = x^TPx + q\int_{\sigma}^{\sigma} f(\sigma)\, d\sigma$, where P is given by $A^TP + AP = -C$, C being a positive definite matrix. If $f(\sigma)$ is constrained as above, then its maximum and minimum values are $K_2\sigma$ and $K_1\sigma$ respectively. One can write,

$$V \leq x^TPx + q\int_{\sigma}^{\sigma} K_2\sigma d\sigma$$

$$\leq x^TPx + \frac{q}{2}K_2\sigma^2.$$

$$V \leq x^TPx + \frac{q}{2}K_2||c||^2\ ||x||^2.$$

Let Ω be the maximum eigenvalue of P. Let λ be the minimum eigenvalue of P. Then,

$$\lambda||x||^2 \leq x^TPx \leq \Omega||x||^2.$$

But,

$$\sup_{||x||=\alpha} V \leq \Omega\alpha^2 + \frac{q}{2} K_2 ||c||^2 \alpha^2.$$

and

$$\inf_{||x||=\beta} V \geq \lambda||x||^2 + \frac{q}{2} K_1 ||x||^2 ||c||^2$$

$$\geq \lambda\beta^2 + \frac{q}{2} K_1 \beta^2 ||c||^2$$

Thus, application of Theorem 9 results in the following:

$$\Omega\alpha^2 + \frac{q}{2} K_2 ||c||^2 \leq \lambda\beta^2 + \frac{q}{2} K_1 ||c||^2 \beta^2$$

$$[\Omega + \frac{q}{2} K_2 ||c||^2]\alpha^2 \leq [\lambda + \frac{q}{2} K_1 ||c||^2]\beta^2$$

When the above condition along with the condition

$$\dot{V}_{(1)} < 0$$

is satisfied the system will be uniformly practically stable.

Application To Transformation Matrix

The object here is to define the error matrix propagation, apply the practical stability concept and obtain the error bound for the given system. As is well known [1] the direction cosine matrix B(t) can be expressed in terms of the four parameter vector E(t). The direction cosine matrix propagates as

$$\dot{B}(t) = \Lambda(t)\ B(t) \tag{33}$$

where $\Lambda(t)$ is a skew symmetric matrix of body angular rates as measured by the system gyroscopes and is given by

$$\Lambda(t) = \begin{bmatrix} 0 & \dot{\phi}_z(t) & -\dot{\phi}_y(t) \\ -\dot{\phi}_z(t) & 0 & \dot{\phi}_x(t) \\ \dot{\phi}_y(t) & -\dot{\phi}_x(t) & 0 \end{bmatrix}$$

$\dot{\phi}_x$, $\dot{\phi}_y$, $\dot{\phi}_z$, are the rotational rates about the vehicle co-ordinate system. Since the actual on-board integration is done in a discrete fashion the true direction cosine matrix propagates as

$$B_{k+1} = \chi_k \, B_k \tag{34}$$

where χ_k is the state transition matrix between states k and k + 1. It is shown in [9] that the series expansion of $\chi_k = e^{\dot{A}_k}$ (the state transition matrix) to a certain fixed number of terms is equivalent to evaluating the equation (33) by the standard numerical integration techniques. In particular, it is shown that the Euler's, Modified Euler's and Runge Kutta (fourth order) methods are equivalent to the first two, three and five terms in the series expansion for e^{A_k} respectively.

$$e^{A_k} = I + A_k + \frac{A_k^2}{2!} + \frac{A_k^3}{3!} + \ldots. \tag{35}$$

Using only the first two terms in (35) will be equivalent to solving equation (33) by Euler's method. Thus the use of an integration algorithm introduces errors and the approximate cosine matrix propagates as

$$\hat{B}_{k+1} = \hat{\chi}_k \, \hat{B}_k \tag{36}$$

here

$$\hat{\chi}_k = I + A_k \tag{37}$$

Define,

$$\tilde{\chi}_k = \hat{\chi}_k - \chi_k \text{ and } E_n = \hat{B}_n - B_n$$

Therefore

$$\tilde{\chi}_k = I + A_k - e^{A_k}.$$

Subtracting (34) from (36) yields

$$E_{k+1} = \hat{\chi}_k \hat{B}_k - \chi_k B_k$$

$$= \hat{\chi}_k \hat{B}_k - \tilde{\chi}_k B_k - \hat{\chi}_k B_k$$

$$= \hat{\chi}_k E_k + \hat{x}_k B_k. \tag{38}$$

Define the z matrix (propagation of error matrix) as

$$Z_{k+1} = B_{k+1}^T E_{k+1}$$

$$= B_k^T \chi_k^T \chi_k B_k Z_k + B_k^T \chi_k^T \tilde{\chi}_k B_k Z_k + B_k^T \chi_k^T \tilde{\chi}_k B_k \tag{39}$$

B and χ_k are both orthogonal matrices in the equations above. Define,

$$R^* = B_k^T \chi_k^T \tilde{\chi}_k B_k . \tag{40}$$

Then the equation for z matrix reduces to

$$Z_{k+1} - Z_k = \Delta Z = R^* Z_k + R^*. \tag{41}$$

If h is the numerical integration step size, then,

$$\frac{\Delta Z}{h} = \frac{1}{h} [R^* Z + R^*]. \tag{42}$$

Using the approximation

$$\dot{Z} \cong \frac{\Delta Z}{h}$$

$$\dot{Z} = RZ + R \tag{43}^{\dagger}$$

$$= R[Z + I] \qquad \text{with } Z(0) = 0 \text{ and } R = \frac{R^*}{h}$$

Let

$$Z + I = Y$$

Therefore,

$$\dot{Z} = \dot{Y} \quad \text{and} \quad Y(0) = I.$$

---

† For details refer to [10].

Hence

$$\dot{Y} = RY \tag{44}$$

is the formulation of the error equation.

The system (44) can be tested for practical stability and if it is practically stable the error bounds can be obtained by picking up a proper V function and obtaining the bounds by the hypothesis (ii) of Theorem 7. In this case the set $||x|| < \alpha$ is given by $||y|| = 1$ and $||y|| = \beta$ gives the bound in y from which the error bound on Z can be easily calculated. The bounds so obtained will be the upper bounds on the error.

VI. CONCLUSIONS

The errors involved in the numerical integration of differential equations are shown to be dependent upon integration step size, mode of operation, word length limitation of the computing machine, etc. Assuming that the round-off errors are random in nature, the analysis is carried out with and without additive noise. It is observed that the noise affects those errors which have a comparable level to the noise input. Future work in this area can be done to correlate the output noise levels to the input noise levels, to determine the effect of noise on optimum step size and to obtain a better estimation of the output with noisy inputs.

It is shown that for the Trapezoidal integration scheme a change of input level changes the errors in both the fixed point and floating point mode of arithmetic operations in the same manner. It is shown that as the input magnitude is increased the optimum step size will decrease but the magnitudes of the errors at the optimum step size will increase.

A method based on the concept of "Practical Stability" is used for developing output bounds for a general class of systems. For a meaningful quantitative theory the system stability is defined in terms of subsets of the state space which are pre-specified in a given problem and which in general may be time varying. The differential equation

to be integrated is converted to an error equation. It is shown that the Practical Stability technique can be applied to this equation to obtain the error bounds of the system. The bounds so obtained will be the upper bounds on the error. This approach is also applied to several examples.

REFERENCES

1. Burdeshaw, D. H., "Methods of computing the transformation matrix associated with gimballess inertial measurement units," NASA Technical memorandum, NASA TM X-53294, Marshall Space Flight Center, Huntsville, Alabama, July 13, 1965.

2. "A study of the critical computational problems associated with strapdown inertial nagiation system," United Aircraft Corporation, NASA CR-968, pp. IV-/ -IV-9, April, 1968.

3. Henrici, Peter, Elements of Numerical Analysis, John Wiley & Sons, pp. 306, 1964.

4. Kamke, E., "Zur theorie der systeme gewöhnlicher differential gleichungen-II, Acta Math. 58(1932) p. 57-85.

5. Coppel, W. A., Stability and Asymptotic Behavior of Differential Equations. Boston: Heath and Company, 1965.

6. Michel, A. N., "Quantitative analysis of simple and interconnected systems: stability, boundedness, and trajectory behavior," IEEE Transactions on Circuit Theory, Volume CT-17, pp. 292-301, August, 1970.

7. Dorato, P. "Short time stability in linear time-varying systems," IRE International Convention Record, Part 4, pp. 83-87, 1961.

8. Cesari, L. "Un nuovo criterio di stabililita per le soluzioni delle equazioni differenziali lineari," Ann. Scuola Norm. Supulisa (2) 9(1940), p. 163-186.

9. Marshall, M. B. and Capehart, B. L., "Numerical solution of state equations," Proceedings IEEE, Vol. 57, pp. 1239-1240, June, 1969.

10. Jordan, J. W., "Direction cosine computational error," NASA Technical Report NASA TR R-304, March, 1969.

2